

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
16 May 2002 (16.05.2002)

PCT

(10) International Publication Number  
WO 02/39371 A2

- (51) International Patent Classification<sup>7</sup>: G06K 9/00 (74) Agent: GROENENDAAL, Antonius, W., M.; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).
- (21) International Application Number: PCT/EP01/12346
- (22) International Filing Date: 23 October 2001 (23.10.2001) (81) Designated State (national): JP.
- (25) Filing Language: English (84) Designated States (regional): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).
- (26) Publication Language: English
- (30) Priority Data: 09/705,666 3 November 2000 (03.11.2000) US Published:  
— without international search report and to be republished upon receipt of that report
- (71) Applicant: KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL). For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.
- (72) Inventors: COLMENAREZ, Antonio, J.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). GUTTA, Srinivas; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).



WO 02/39371 A2

(54) Title: ESTIMATION OF FACIAL EXPRESSION INTENSITY USING A BIDIRECTIONAL STAR TOPOLOGY HIDDEN MARKOV MODEL.

(57) Abstract: An image processing system processes a sequence of images using a bidirectional star topology hidden Markov model (HMM) to estimate facial expression intensity or other characteristics. The HMM has at least one neutral expression state and a plurality of expression paths emanating from the neutral expression state. Each of the expression paths includes a number of states associated with a corresponding facial expression, such as sad, happy, anger, fear, disgust and surprise. A given expression path may include an initial state coupled to the neutral state and a final state associated with an apex of the corresponding expression. The given path further includes a forward path from the initial state to the final state, associated with an onset of the expression, and a return path from the final state to the initial state, associated with an offset of the expression. Control of one or more actions in the image processing system may be based at least in part on which of the facial expressions supported by the model is determined to be present in the sequence of images and/or the intensity or other characteristic of that facial expression.

Estimation of facial expression intensity using a bidirectional star topology hidden markov model

#### Field of the Invention

The present invention relates generally to the field of image signal processing, and more particularly to techniques for estimating facial expression in a video signal or other type of image signal.

5

#### Background of the Invention

Facial expressions have been widely studied from psychological and computer vision points of view. Such expressions provide a mechanism to show emotions, which are crucial in inter-personal communications, relationships, and many other contexts. A number of different types of facial expressions have been determined to be consistent across most races and cultures. For example, certain distinct facial expressions are associated with emotional states. These include neutral, happiness, sadness, anger and fear. Other facial expressions are associated with reactions such as disgust and surprise.

It is known that facial expressions are complex, spatio-temporal motion patterns. The movements associated with a given facial expression are generally divided into three periods: (i) onset, (ii) apex, and (iii) offset. These periods correspond to the transition towards the facial expression, the period sustaining the peak in expressiveness, and the transition back from the expression, respectively. The rate of change in the onset period as well as the duration in the apex period are often related to the intensity of the underlying emotion associated with the facial expression. Similarly, there is evidence that differences in speed during the onset and offset periods can be used to discriminate between spontaneous and fake facial expressions.

Early computer vision-based expression recognition algorithms generally relied solely on the apex of the facial deformation, discarding most of the spatio-temporal information present in the transitions. More recently developed techniques attempt to exploit the spatio-temporal information. For example, facial expression recognition techniques based on hidden Markov models (HMMs), as described in T. Otsuka et al., "Recognizing Abruptly Changing Facial Expressions From Time-Sequential Face Images," International Conference on Computer Vision and Pattern Recognition (CVPR), 1998; and T. Otsuka et al.,

"Recognizing Multiple Persons' Facial Expression Using HMM Based on Automatic Extraction of Significant Frames from Image Sequences," International Conference on Image Processing (ICIP), pp. 546-549, 1997, are specifically designed to take advantage of the spatio-temporal character of facial expression patterns.

5                Nonetheless, most conventional facial expression recognition approaches are based primarily on the analysis of non-rigid facial deformation patterns. These approaches may utilize techniques such as optical flow, two-dimensional graphical models (also known as "potential nets") and local parametric models. Unfortunately, facial appearance changes due to expression deformations are not always well described by motion fields. For example,  
10        exposing the teeth as in a smile is not well represented by the motion field of the mouth area.

                 A Bayesian framework for embedded face and facial expression recognition which overcomes the above-noted problems is described in A. Colmenarez et al., "A Probabilistic Framework for Embedded Face and Facial Expression Recognition,"  
15        International Conference on Computer Vision and Pattern Recognition (CVPR), 1999; A. Colmenarez et al., "Embedded Face and Facial Expression Recognition," International Conference on Image Processing (ICIP), 1999; A. Colmenarez et al., "Detection and Tracking of Faces and Facial Features," International Conference on Image Processing (ICIP), 1999; and A. Colmenarez, "Facial Analysis from Continuous Video with Application  
20        to Human-Computer Interface," Ph.D. dissertation, University of Illinois at Urbana-Champaign, March 1999, all of which are incorporated by reference herein. By modeling and analyzing the appearance and geometry of facial features under different facial expressions for different people, the above-noted Bayesian framework is able to achieve both face recognition and facial expression recognition. However, the framework generally does not  
25        take into account the dynamics of the facial expressions. For example, this approach generally assumes that image frames from a given video signal to be analyzed are independent of each other, and therefore analyzes them one frame at a time.

                 A need therefore exists for improved techniques for estimating facial expression, in a manner that accurately and efficiently takes into account facial expression  
30        dynamics.

#### Summary of the Invention

                 The invention provides methods and apparatus for processing a video signal or other type of image signal in order to estimate facial expression intensity or other

characteristics associated with facial expression dynamics, using a bidirectional star topology hidden Markov model (HMM).

In accordance with one aspect of the invention, the HMM has at least one neutral expression state and a plurality of expression paths emanating from the neutral expression state. Each of the expression paths includes a number of states associated with a  
5 corresponding facial expression, such as sad, happy, anger, fear, disgust and surprise. Control of one or more actions in the image processing system may be based at least in part on which of the facial expressions supported by the model is determined to be present in the sequence of images and/or the intensity or other characteristic of that expression.

10 In accordance with another aspect of the invention, a given expression path of the HMM may include an initial state coupled to the neutral state and a final state associated with an apex of the corresponding expression. The given path further includes a forward path from the initial state to the final state, and a return path from the final state to the initial state. The forward path is associated with an onset of the expression, and the return path is  
15 associated with an offset of the expression. The forward and reverse paths of the given expression path may each include separate states, or may share a number of states.

In accordance with a further aspect of the invention, each of at least a subset of the states of a given one of the expression paths may be interconnected in the HMM with at least one state of at least one other expression path by an interconnection which does not pass  
20 through the neutral state.

The invention provides significantly improved estimation of facial expression relative to the conventional techniques described previously. Advantageously, the invention allows one to determine not only the particular facial expression present within a given image, but also the intensity or other relevant characteristics of that facial expression. The  
25 techniques of the invention can be used in a wide variety of image processing applications, including video-camera-based systems such as video conferencing systems, video surveillance and monitoring systems, and human-machine interfaces.

#### Brief Description of the Drawings

30 FIG. 1 is a block diagram of an image processing system in which the present invention may be implemented.

FIG. 2 shows an example of a model of facial features and regions that may be used in conjunction with estimation of facial expression in an illustrative embodiment of the invention.

FIG. 3 shows an example of a Bayesian network for observation distribution suitable for use in the illustrative embodiment of the invention.

FIG. 4 shows an example of a bidirectional star topology hidden Markov model (HMM) configured in accordance with the invention.

5           FIGS. 5 and 6 show alternative configurations for the bidirectional star topology HMM of FIG. 4 in accordance with the invention.

#### Detailed Description of the Invention

FIG. 1 shows an image processing system 10 in which facial expression  
10 estimation techniques in accordance with the invention may be implemented. The system 10 includes a processor 12, a memory 14, an input/output (I/O) device 15 and a controller 16, all of which are connected to communicate over a set 17 of one or more system buses or other type of interconnections. The system 10 further includes a camera 18 that is coupled to the controller 16 as shown. The camera 18 may be, e.g., a mechanical pan-tilt-zoom (PTZ)  
15 camera, a wide-angle electronic zoom camera, or any other suitable type of image capture device. It should therefore be understood that the term "camera" as used herein is intended to include any type of image capture device or any configuration of multiple such devices.

The system 10 may be adapted for use in any of a number of different image processing applications, including, e.g., video conferencing, video surveillance, human-  
20 machine interfaces, etc. More generally, the system 10 can be used in any application that can benefit from the improved facial expression estimation capabilities provided by the present invention.

In operation, the image processing system 10 generates a video signal or other type of sequence of images of a person 20. The camera 18 may be adjusted such that a head  
25 24 of the person 20 comes within a field of view 22 of the camera 18. A video signal corresponding to a sequence of images generated by the camera 18 and including a face of the person 20 is then processed in system 10 using the facial expression estimation techniques of the invention, as will be described in greater detail below. The sequence of images may be processed so as to determine a particular expression that is on the face of the  
30 person 20 within the images, based at least in part on an estimation of the intensity or other characteristic of the expression as determined using a bidirectional star topology hidden Markov model (HMM). An output of the system may then be adjusted based on the determined expression. For example, a human-machine interface or other type of system application may generate a query or other output or take another type of action based on the

determined expression or characteristic thereof. Any other type of control of an action of the system may be based at least in part on the determined expression and/or a particular characteristic thereof, such as intensity.

Elements or groups of elements of the system 10 may represent corresponding  
5 elements of an otherwise conventional desktop or portable computer, as well as portions or combinations of these and other processing devices. Moreover, in other embodiments of the invention, some or all of the functions of the processor 12, memory 14, controller 16 and/or other elements of the system 10 may be combined into a single device. For example, one or more of the elements of system 10 may be implemented as an application specific integrated  
10 circuit (ASIC) or circuit card to be incorporated into a computer, television, set-top box or other processing device.

The term "processor" as used herein is intended to include a microprocessor, central processing unit (CPU), microcontroller, digital signal processor (DSP) or any other data processing element that may be utilized in a given data processing device. In addition, it  
15 should be noted that the memory 14 may represent an electronic memory, an optical or magnetic disk-based memory, a tape-based memory, as well as combinations or portions of these and other types of storage devices.

The present invention in an illustrative embodiment provides techniques for estimating facial expression in an image signal, and for characterizing dynamic aspects of  
20 facial expression using an HMM. The invention in the illustrative embodiment models transitions between different facial expressions as well as transitions between multiple states within each facial expression. More particularly, each expression is modeled as multiple states along a path in a multi-dimensional space of facial appearance. This path for a given expression goes from a point corresponding to a neutral expression to that of an apex of the  
25 expression and back to the neutral expression.

Advantageously, the invention allows one to determine not only the particular facial expression present within a given image, but also the intensity or other relevant characteristic of that facial expression. The former may be obtained using maximum likelihood classification among a set of different facial expression models. The latter may be  
30 estimated by determining how far the observation reaches in terms of the above-noted path of the corresponding facial expression.

An exemplary facial expression analysis framework suitable for use in conjunction with the present invention will now be described in greater detail. Consider a

framework in which  $\rho \in \{1, 2, \dots, P\}$  is an index to a  $\rho$ -th person in a database of  $P$  people and  $V$  is a portion of a video signal used for facial analysis, e.g., face and facial expression recognition. Face recognition can be carried out using maximum likelihood,

$$\rho^* = \arg \max_{\rho=1, \dots, P} P(V|\rho), \quad (1)$$

by selecting the model that maximizes the likelihood probability of the observed image. It should be noted, however, that the invention does not require that a given image or sequence of images be identified as being associated with a particular person.

Once the person  $\rho$  is recognized, a corresponding person-dependent facial model may be used to carry out facial expression recognition. An example of such a model will be described below in conjunction with FIG. 2.

In accordance with the invention, an HMM is used to capture the dynamics of facial expressions. More particularly, in the illustrative embodiment of the invention, hidden states  $\epsilon = 1, 2, \dots, N$  of an HMM are used to represent different facial expression stages, and transition probability matrices  $P(\epsilon_t | \epsilon_{t-1}, \rho)$  capture the statistics of the facial expression dynamics for a given person  $\rho$ .

Consider a video segment  $V_t = \{f_1, f_2, \dots, f_t\}$  as a collection of sequential image frames up to time  $t$ . The likelihood probability of this video segment for a given person  $\rho$ ,  $P(V_t | \rho)$ , may be computed recursively from

$$P(V_t | \rho) = P(V_{t-1} | \rho) P(\epsilon_t | \epsilon_{t-1}, \rho) P(f_t | \epsilon_t, \rho), \quad (2)$$

where  $P(f_t | \epsilon_t, \rho)$  is the observation probability for a given expression state  $\epsilon$  and person  $\rho$ .

FIG. 2 shows a model of facial features and regions that may be used in conjunction with the estimation of facial expression in the illustrative embodiment of the invention. It should be understood that this model is provided by way of example only, and should not be construed as limiting the scope of the invention in any way. As will be apparent to those skilled in the art, the invention can be implemented using a wide variety of other facial models.

In the facial model of FIG. 2, the face of a person in an image 40 is modeled as a set of four feature regions and nine facial features. The position of each of the facial features in this example is denoted by an X. The four feature regions include a right eyebrow region 50-1, a left eyebrow region 50-2, an eyes and nose region 50-3, and a mouth region 50-4.

This model is constructed under the assumption that the facial features can be accurately located and tracked in video sequences, as is described in greater detail in the above-cited reference A. Colmenarez et al., "Detection and Tracking of Faces and Facial Features," ICIP, 1999.

The appearance of each facial feature is provided by a corresponding feature image sub-window located around its position. More particularly, the right eyebrow region 50-1 includes image sub-windows 52 and 54, the left eyebrow region 50-2 includes image sub-windows 62 and 64, the eyes and nose region 50-3 includes image sub-windows 72, 74 and 76, and the mouth region 50-4 includes image sub-windows 82 and 84. Facial geometry is given by the facial feature positions, which may be normalized with respect to the position and distance between the outer eye corners.

One can assume that the facial feature regions  $\{\mathbf{r}_i \mid i = 1, 2, \dots, R\}$  are independent for a given person and facial expression state, and then compute the likelihood probability of the observed image frame from

$$\mathbf{P}(\mathbf{f}|\mathbf{e}, \rho) = \prod_{k=1}^R \mathbf{P}(\mathbf{r}_k|\mathbf{e}, \rho). \quad (3)$$

On each region, the likelihood probability  $\mathbf{P}(\mathbf{r}_k|\mathbf{e}, \rho)$  may be computed using the positions  $\mathbf{x}_{ki}$  and appearances  $\mathbf{v}_{ki}$  of its  $F_k$  features as:

$$\begin{aligned} \mathbf{P}(\mathbf{r}_k|\mathbf{e}, \rho) = & \\ & \mathbf{P}(\mathbf{v}_{k1}, \dots, \mathbf{v}_{kF_k} | \mathbf{x}_{k1}, \dots, \mathbf{x}_{kF_k}, \mathbf{e}, \rho) \\ & \mathbf{P}(\mathbf{x}_{k1}, \dots, \mathbf{x}_{kF_k} | \mathbf{e}, \rho) \end{aligned} \quad (4)$$

The position of the facial features in a region  $\mathbf{P}(\mathbf{x}_{k1}, \dots, \mathbf{x}_{kF_k} | \mathbf{e}, \rho)$  may be modeled jointly with a multi-dimensional Gaussian distribution having a full-covariance



matrix. In addition, the appearance of each facial feature in a region may be modeled independently, so that Equation (4) becomes

$$\begin{aligned} P(\mathbf{r}_k | \varepsilon, \rho) = & \\ P(\mathbf{v}_{ki} | \mathbf{x}_{ki}, \varepsilon, \rho) \dots P(\mathbf{v}_{kFk} | \mathbf{x}_{kFk}, \varepsilon, \rho) & \quad (5) \\ P(\mathbf{x}_{k1}, \dots, \mathbf{x}_{kFk} | \varepsilon, \rho) & \end{aligned}$$

5

FIG. 3 shows an example of a Bayesian network that may be used to model facial appearance and geometry observations in the illustrative embodiment of the invention. Associated with each state  $k$  of a given observed facial expression are a set of feature regions 50-1, 50-2, 50-3 and 50-4 and their corresponding set of feature positions 100 and feature image sub-windows 110.

The appearance of each facial feature for a given person and expression state may be modeled with a multi-dimensional Gaussian distribution applied over the  $p$  principal components and the distance from this sub-space. Note that this approach is different from a conventional eigenfeatures approach, in which Principal Component Analysis (PCA) is used to find the sub-space in which all object classes span the most. This approach is different in that PCA is applied to each class independently in order to construct simple observation models that handle the high dimensionality of the observations.

Let  $\mathbf{v} \in \mathbb{R}^d$  be a  $d$ -dimensional random vector with some distribution that is to be modeled, i.e., an image sub-window around the corresponding facial feature position. A set of training samples of a class is used to estimate the mean  $\bar{\mathbf{v}}$  and the covariance matrix  $\Omega$  of the observation of that class. Using singular value decomposition, one can obtain the diagonal matrix  $\Sigma$  corresponding to the  $p$  largest eigenvalues of  $\Omega$ , and the transformation matrix  $T$  containing the corresponding eigenvectors. So, the conditional probability of  $\mathbf{v}$  for a given class is computed from

25

$$\begin{aligned} P(\mathbf{v}) = & \\ \frac{1}{\sqrt{(2\pi)^p \det(\Sigma)}} \exp\left[-\frac{1}{2} \mathbf{u}' \Sigma^{-1} \mathbf{u}\right] & \quad (6) \\ \frac{1}{\sqrt{2\pi\lambda}} \exp\left[-\frac{d^2}{2\lambda}\right] & \end{aligned}$$

where  $\mathbf{u} = \mathbf{T}(\mathbf{v} - \bar{\mathbf{v}})$  is the projection of  $\mathbf{v}$  onto the above-noted  $p$ -dimensional subspace,  $d = \sqrt{|\mathbf{v} - \bar{\mathbf{v}}|^2 - |\mathbf{u}|^2}$  is the distance from this subspace, and  $\lambda$  is obtained from the sum of the remainder of the eigenvalues of  $\Omega$ .

Note that the above-described learning procedure is supervised. In other words, it assumes that the class of each training sample is known so that the statistics for each class can be easily computed. In the case of the above-noted facial expression database, video segments may be labeled so that the facial expression is known for each image frame. However, as each facial expression is modeled with multiple states that are to capture different stages of the facial expression evolution over time, the training procedure of these states within each facial expression is unsupervised. A conventional Expectation-Maximization (EM) algorithm may then be used to estimate the parameters of the observation model as set forth in equations (3) to (6). The EM algorithm is described in greater detail in, e.g., B.J. Frey, "Graphical Models for Machine Learning and Digital Communication," MIT Press, Cambridge, MA, 1998, which is incorporated by reference herein.

The modeling of expression dynamics in accordance with the invention will now be described in greater detail, with reference to FIGS. 4, 5 and 6. Consider a multi-dimensional space of facial appearance and geometry. A point in this space corresponds to a particular observation of a face. Therefore, one can define a neutral-expression point as the point in this multi-dimensional space corresponding to the neutral expression for a given face. Similarly, one can define other points corresponding to other facial expressions. During an onset period of a given facial expression, as the face in question changes to reflect the given facial expression, a path is followed in the above-noted space, i.e., from the neutral-expression point to the apex of the corresponding expression. Similarly, another (or possibly the same) path is followed back from the apex to the neutral-expression point in the offset period. Other facial expressions produce different paths, and the different paths may cross one another.

FIG. 4 shows an example of a set of the above-noted paths in the form of a bidirectional star topology HMM 120 having a single neutral expression state 122. From the neutral expression state 122 there are a set of six different paths, each corresponding to a particular facial expression and each including a total of  $N$  states. In this example, the facial

expressions modeled are happy, sad, anger, fear, disgust and surprise. Other embodiments could of course use other numbers, types and arrangements of facial expressions. For simplicity, each expression is modeled with a path of multiple interconnected states. In addition, it is assumed that forward and return paths, which correspond to the onset and offset periods, respectively, are the same. The neutral expression is modeled with a single state, the neutral expression state 122. The neutral state 122 is connected to the first state of each facial expression path. In the case of an expression observation that reaches the highest expression intensity modeled, all of the states of the path are visited, first in forward order and then in backward order, returning to the neutral expression state 122.

Assuming that sufficient data is available for training, the bidirectional star topology HMM of FIG. 4 captures the evolution of facial expressions at all levels of intensity. Each state represents one step towards the maximum level of expressiveness associated with the last state in the corresponding facial expression path. During subsequent facial expression analysis, an observation does not necessarily have to reach the last state in the path. Therefore, one can measure the intensity of the observed facial expression using the highest state visited in the path as well as the duration of its visit to that state.

The separation between two consecutive states in the HMM of FIG. 4 can be determined using the well-known Kullback-Leibler divergence of the corresponding observation probability distributions computed along the line that connects the observation points with maximum likelihood for each state. That is,

$$D(s_1, s_2) = \int P(\mathbf{v}|s_1) \log \frac{P(\mathbf{v}|s_1)}{P(\mathbf{v}|s_2)} \delta [\alpha(\bar{\mathbf{v}}_2 - \bar{\mathbf{v}}_1)] \quad (7)$$

where  $\bar{\mathbf{v}}_1$  and  $\bar{\mathbf{v}}_2$  are the mean vectors in the case of Gaussian models,

$P(\mathbf{v}|s_1) = N(\mathbf{v}; \bar{\mathbf{v}}_1, \Omega_1)$  and  $P(\mathbf{v}|s_2) = N(\mathbf{v}; \bar{\mathbf{v}}_2, \Omega_2)$  for the observation probability distributions of states  $s_1$  and  $s_2$ , respectively. It should be noted that this type of divergence is given by way of example only, and other types of divergence could also be used.

It is also important to determine the appropriate number of states for each path. This may be done as follows. Each path is first trained by assuming a default number of states and measuring the average separation between the states. Then, the number of states is

iteratively increased or reduced, and the HMM path is retrained until the average separation is within a predefined range. Additional details on techniques for determining the appropriate number of states in a given path of the HMM 120 may be found in U.S. Patent Application Attorney Docket No. 701255 entitled "Method and Apparatus for Determining a Number of States for a Hidden Markov Model in a Signal Processing System," filed concurrently herewith in the name of inventors A. Colmenarez and S. Gutta, which application is incorporated by reference herein.

Although each path in the FIG. 4 HMM is shown as including the same number of states N, this is by way example and not limitation. In other embodiments, each of the paths may include a different number of states, with the above-described training procedure used to determine the appropriate number of states for a given expression path of the HMM.

Numerous alternative configurations of a bidirectional star topology HMM in accordance with the invention are possible. Examples of two of such alternative configurations will be described below in conjunction with FIGS. 5 and 6. The term "bidirectional star topology HMM" as used herein is intended to include any HMM having one or more bidirectional paths emanating outward from at least one neutral state, and thus includes the alternative arrangements described below as well as numerous other arrangements.

FIG. 5 shows a portion of an alternative bidirectional star topology HMM 120'. The portion shown includes the expression path for the facial expression of surprise. As in the HMM 120 of FIG. 4, the surprise facial expression in HMM 120' starts from the neutral expression state 122. However, the surprise facial expression path in HMM 120' includes two separate paths for modeling the transitions from the neutral state to the expression apex and from that expression apex back to the neutral state. Such an arrangement can be used to provide hysteresis in the state transition process. Although only a single expression is shown in FIG. 5, similar separate paths may be provided for each of the expressions in the HMM.

FIG. 6 shows another alternative configuration of a bidirectional star topology HMM in accordance with the invention. In this configuration, an HMM 120'' includes the neutral state 122 and the expression paths for the same six expressions as the HMM 120 of FIG. 4. Only a portion of each of the expression paths is shown for simplicity of illustration. In the HMM 120'', the first states of each of the different expression paths are interconnected with the first states from one or more other expression paths as shown. For example, the state

Happy 1 is interconnected with the states Disgust 1, Surprise 1 and Sad 1, the state Surprise 1 is interconnected with the states Happy 1, Anger 1 and Fear 1, and so on.

This type of interconnection of some or all of the first few states of each expression path allows transitions from one expression to another without going through the neutral state 122. Note that the interconnection may include adjacent paths in the star topology, such as the happy and surprise paths in the figure, as well as non-adjacent paths, such as the surprise and fear, disgust and happy, and sad and happy paths. As previously noted, the particular states, paths and interconnections shown in FIG. 6 are only an example, and numerous other configurations are possible.

The above-described embodiments of the invention are intended to be illustrative only. For example, the invention can be implemented using a bidirectional star topology HMM having any number or arrangement of expression paths, states and interconnections. The invention can be used to provide facial expression estimation in a wide variety of applications, including video conferencing systems, video surveillance systems, and other camera-based systems. The invention can be implemented at least in part in the form of one or more software programs which are stored on an electronic, magnetic or optical storage medium and executed by a processing device, e.g., by the processor 12 of system 10. These and numerous other embodiments within the scope of the following claims will be apparent to those skilled in the art.

## CLAIMS:

1. A method for use in estimation of facial expression in a sequence of images generated in an image processing system (10), the method comprising the steps of:  
processing the sequence of images using a hidden Markov model (120) having at least one neutral expression state (122) and a plurality of expression paths emanating from  
5 the neutral expression state, each of the expression paths comprising a plurality of states associated with a corresponding facial expression, wherein the processing step utilizes the model to determine a characteristic of a particular one of the facial expressions likely to be present in the sequence of images; and  
controlling an action of the image processing system based on the determined  
10 facial expression characteristic.
2. The method of claim 1 wherein the determined facial expression characteristic comprises an intensity of a particular facial expression.
- 15 3. The method of claim 1 wherein the plurality of expression paths include expression paths corresponding to one or more of the following facial expressions: happy, sad, anger, fear, disgust and surprise.
4. The method of claim 1 wherein each of at least a subset of the expression  
20 paths includes a separate forward path and a separate return path, with each of the forward and return paths including a plurality of states.
5. The method of claim 1 wherein each of at least a subset of the states of a given  
one of the expression paths is interconnected in the hidden Markov model with at least one  
25 state of at least one other expression path by an interconnection which does not pass through the neutral expression state.

6. The method of claim 1 wherein each of at least a subset of the expression paths includes a plurality of states including a first state adjacent to the neutral expression state and a final state associated with an apex of the corresponding facial expression.
- 5 7. The method of claim 6 wherein a forward path through a given one of the expression paths from the first state to the final state is associated with an onset of the corresponding facial expression, and a return path through the given expression path from the final state to the first state is associated with an offset of the corresponding facial expression.
- 10 8. The method of claim 1 wherein the controlling step comprises generating an output of the image processing system based on the determined facial expression characteristic.
9. The method of claim 1 wherein the controlling step comprises altering an  
15 operating parameter of the image processing system based on the determined facial expression characteristic.
10. The method of claim 1 further including the step of processing the sequence of images using the hidden Markov model so as to recognize a face of a particular person within  
20 the sequence of images.
11. The method of claim 1 further wherein each of the expression paths of the hidden Markov model includes the same number of states.
- 25 12. The method of claim 1 further wherein each of at least a subset of the expression paths of the hidden Markov model includes a different number of states.
13. An apparatus for use in estimation of facial expression in a sequence of images generated in an image processing system (10), the apparatus comprising:  
30 a processor-based device (12) operative: (i) to process the sequence of images using a hidden Markov model (120) having at least one neutral expression state (122) and a plurality of expression paths emanating from the neutral expression state, each of the expression paths comprising a plurality of states associated with a corresponding facial expression, wherein the model is utilized to determine a characteristic of a particular one of

the facial expressions likely to be present in the sequence of images; and (ii) to control an action of the image processing system based on the determined facial expression characteristic.

- 5    14.            An article of manufacture comprising a storage medium for storing one or more programs for use in estimation of facial expression in a sequence of images generated in an image processing system (10), wherein the one or more programs when executed by a processor (12) implement the step of:
- processing the sequence of images using a hidden Markov model (120) having
- 10    at least one neutral expression state (122) and a plurality of expression paths emanating from the neutral expression state, each of the expression paths comprising a plurality of states associated with a corresponding facial expression, wherein the processing step determines a characteristic of a particular one of the facial expressions likely to be present in the sequence of images;
- 15                   and further wherein an action of the image processing system is controlled based on the determined facial expression characteristic.



1/6

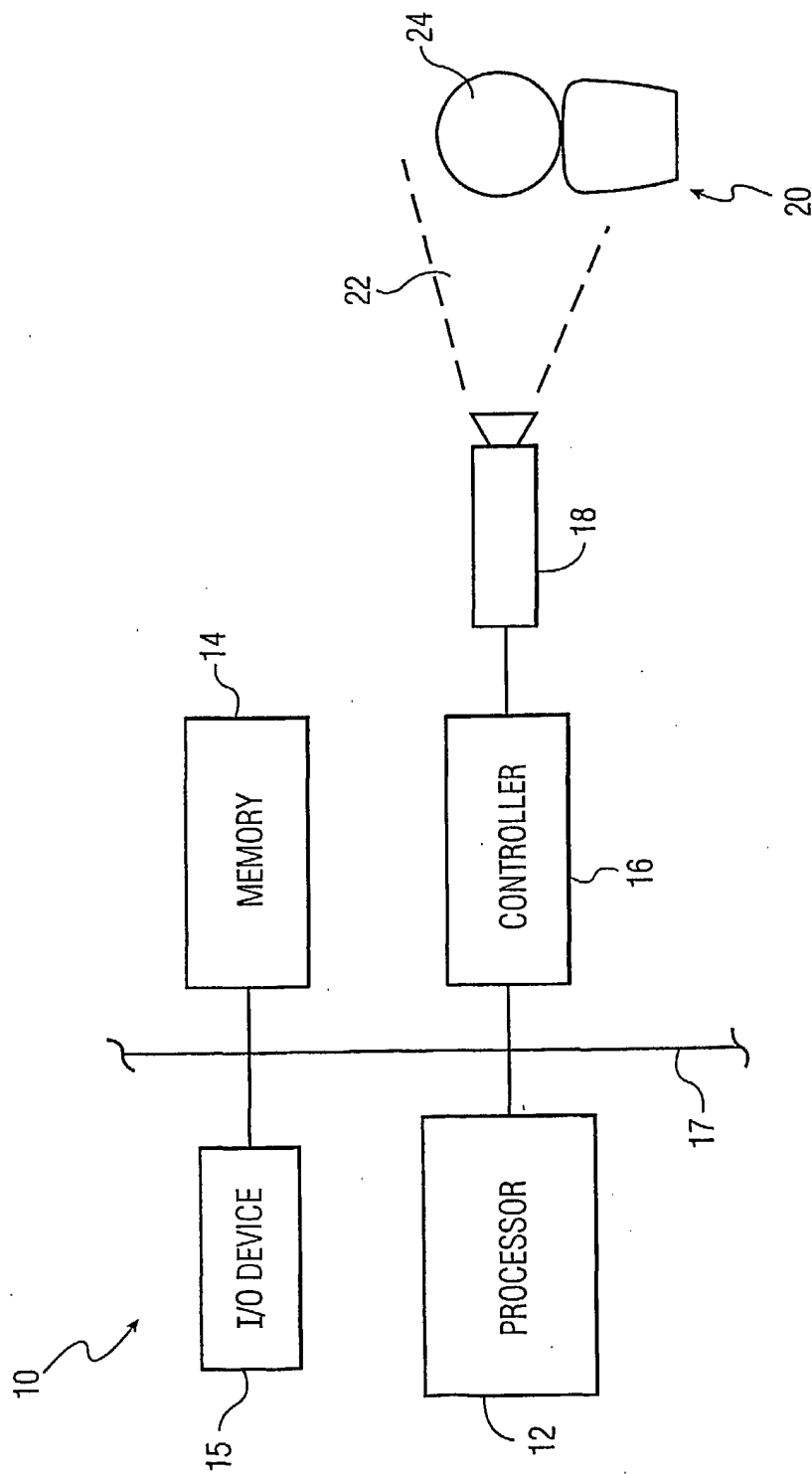


FIG. 1

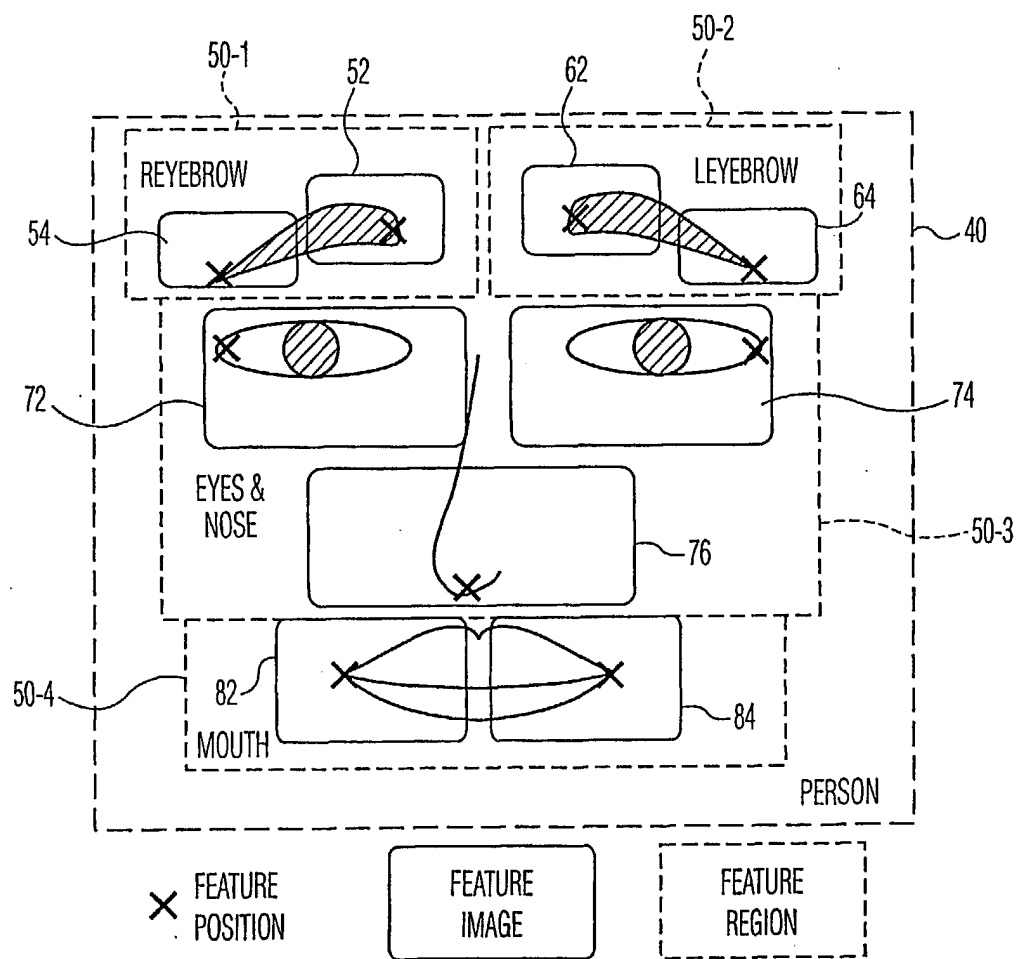


FIG. 2

3/6

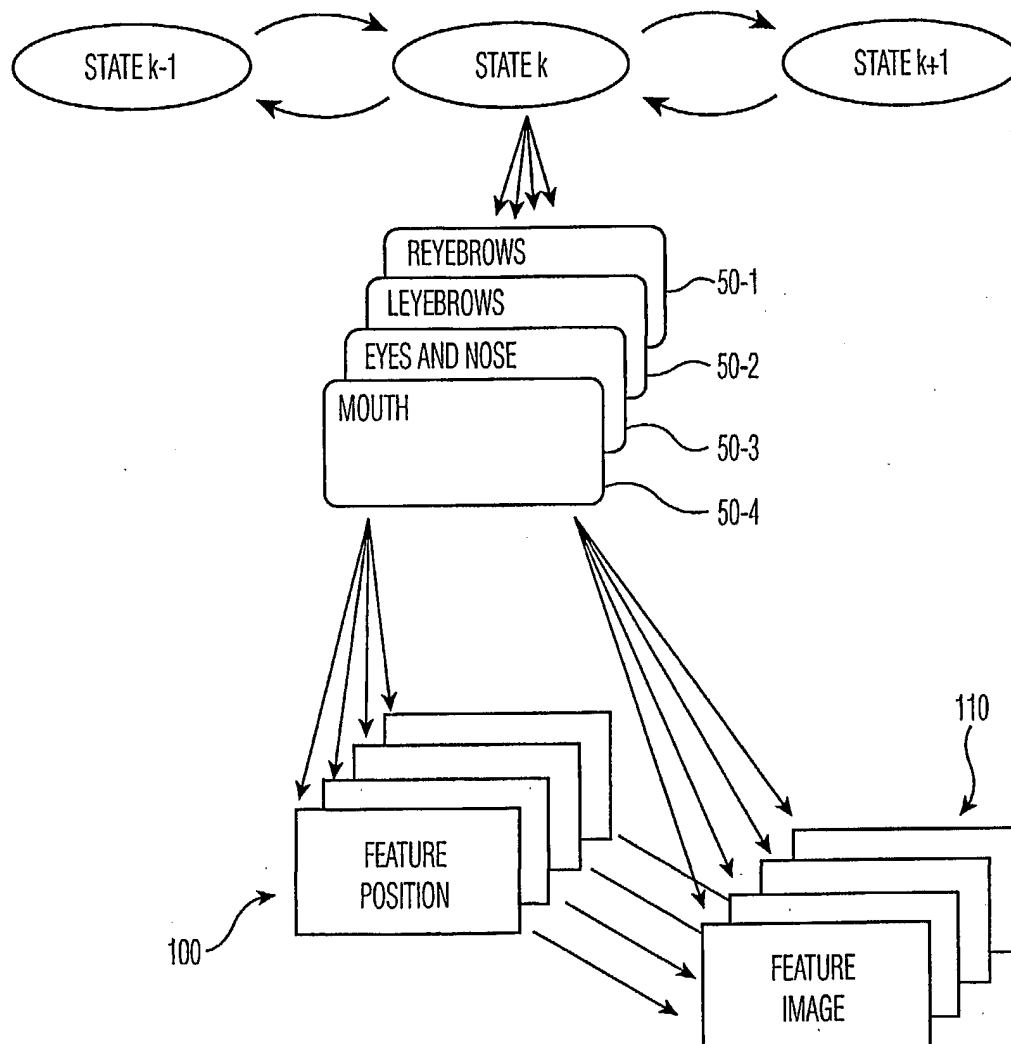


FIG. 3

4/6

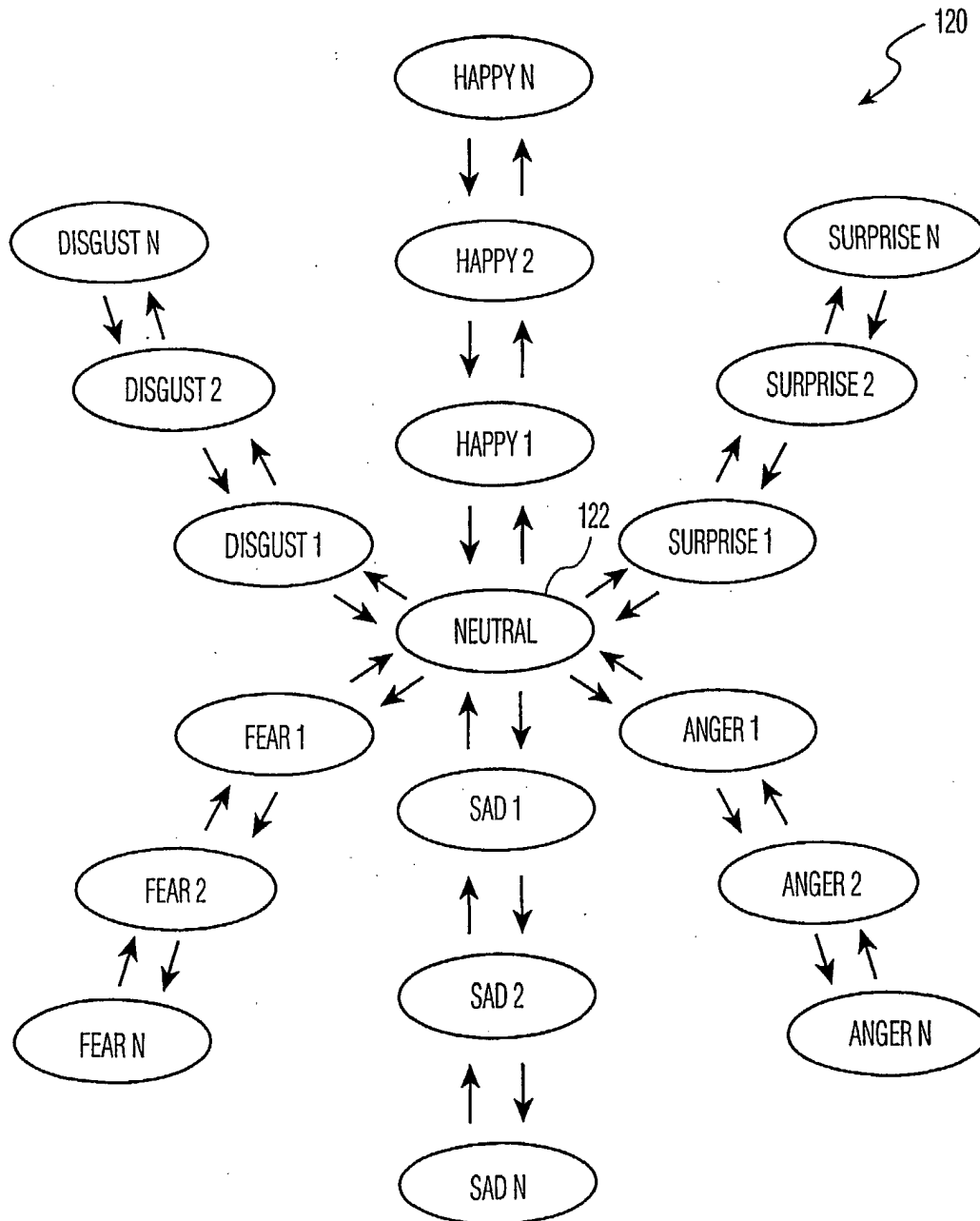


FIG. 4

5/6

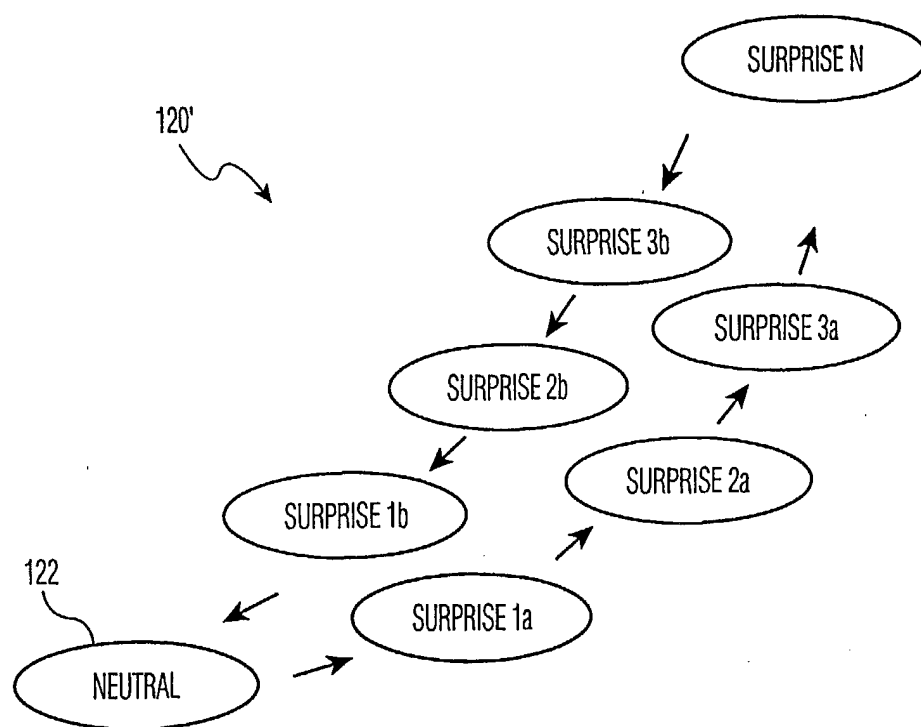


FIG. 5

6/6

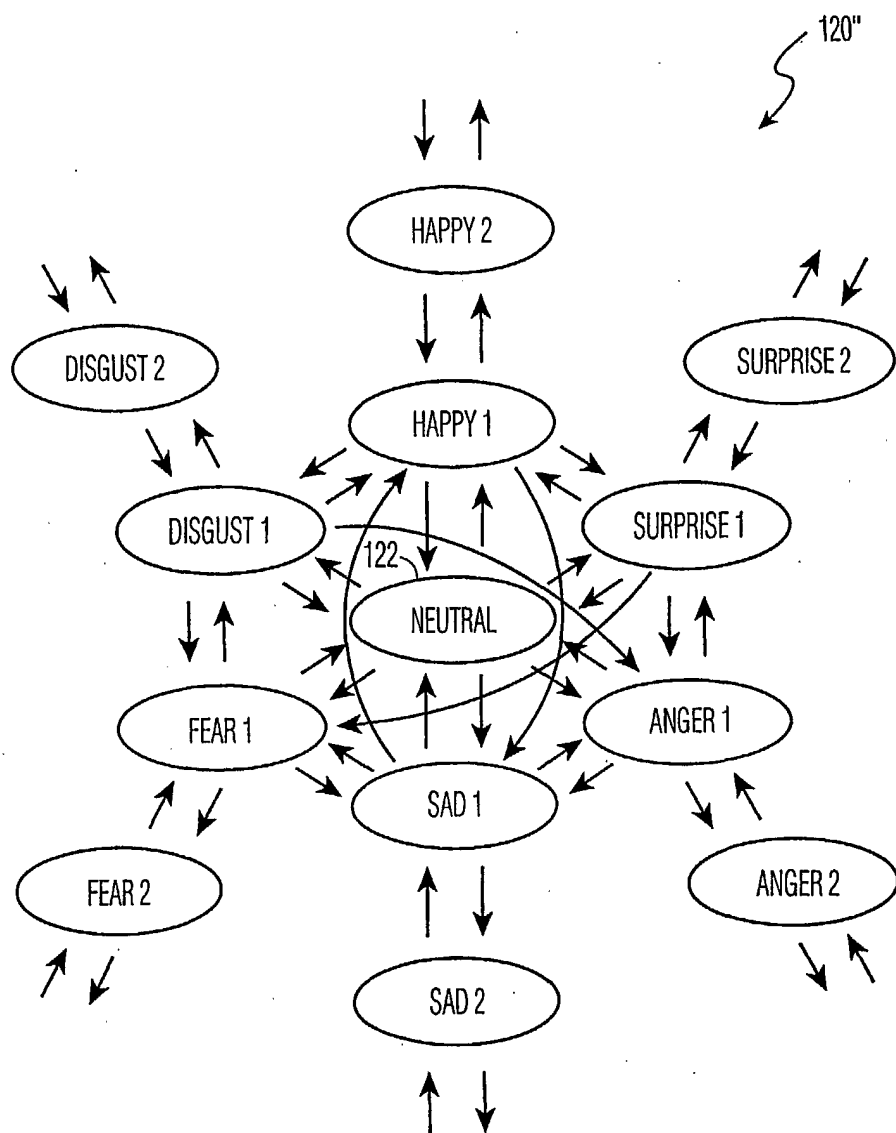


FIG. 6